

AWS

S U M M I T

# Amazon AI 入門

D4T3-1 (AWS Techトラック 3)

パートナー ソリューション アーキテクト  
河原 哲也

2017/6/2 12:20-13:00



WALK OUT  
WITH NO CHECKOUT

**JUST  
WALK  
OUT**

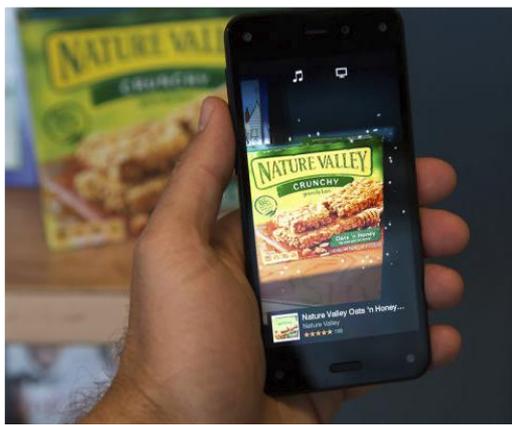
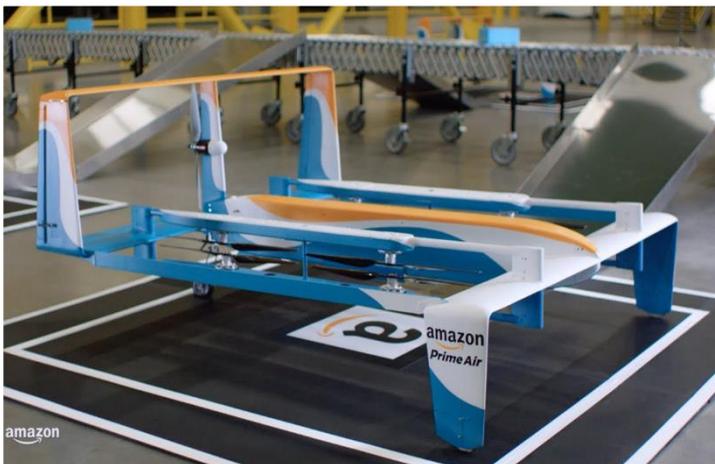
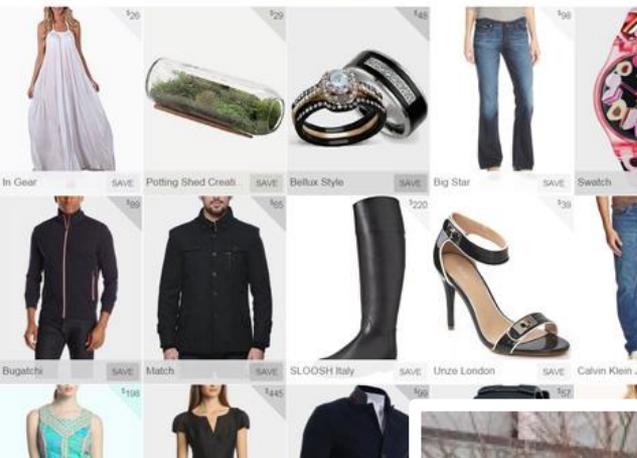
SHOPPING

amazon



# Beautiful things, updated daily

ALL WOMEN MEN YOUR SAVES





# Amazon AI の構成要素

## AI SERVICES



AMAZON  
REKOGNITION  
IMAGE RECOGNITION



AMAZON  
POLLY  
TEXT-TO-SPEECH



AMAZON LEX  
VOICE AND TEXT  
CHATBOTS

## AI PLATFORMS



AMAZON  
MACHINE  
LEARNING



AMAZON  
EMR



SPARK &  
SPARKML

## AI FRAMEWORKS

AWS DEEP LEARNING AMI

APACHE  
MXNET

TENSOR-  
FLOW

CAFFE

TORCH

THEANO

CNTK

KERAS

## AI INFRASTRUCTURE

AMAZON EC2 P2  
AND G2 GPU

AMAZON EC2  
CPUS

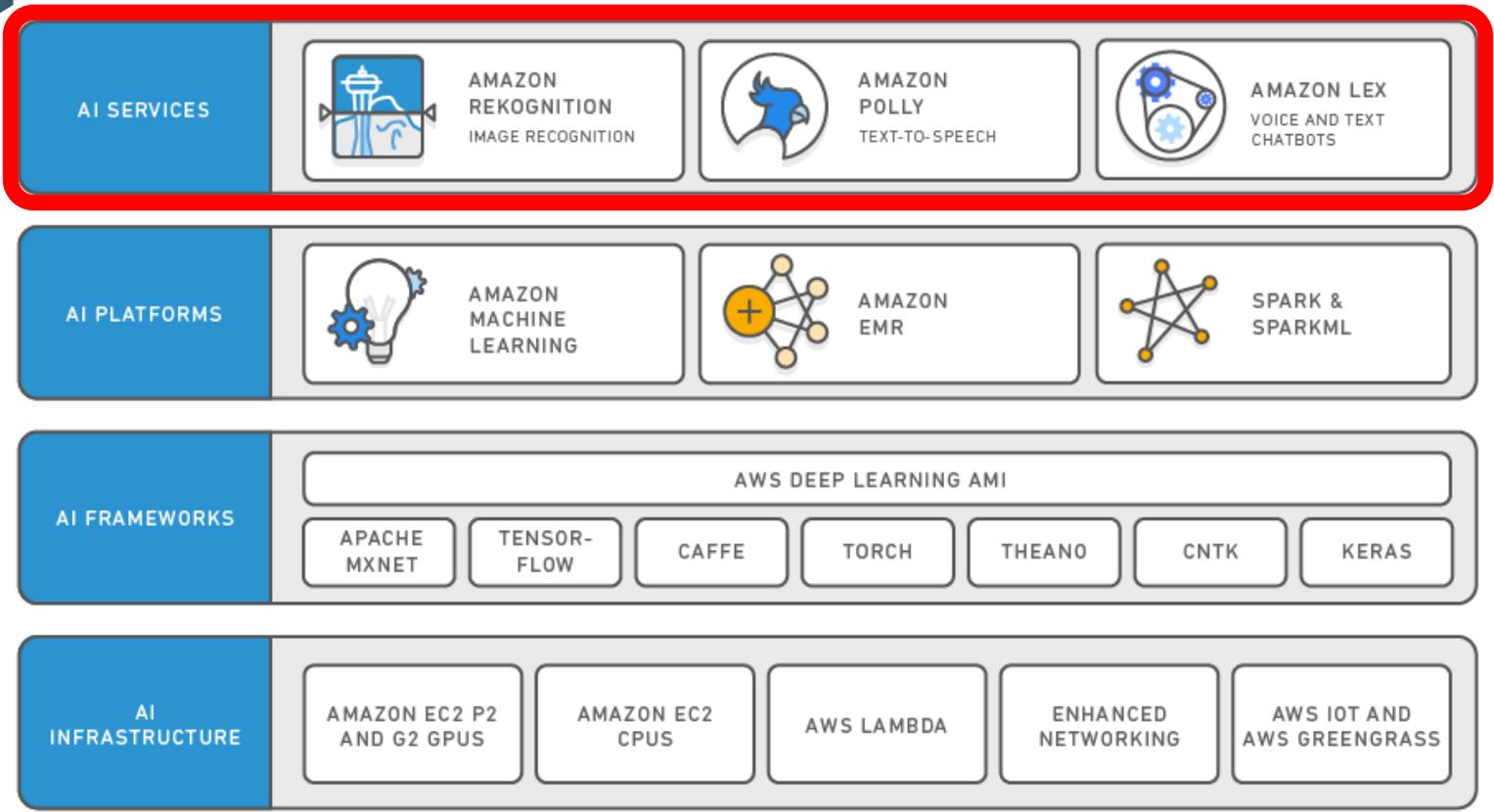
AWS LAMBDA

ENHANCED  
NETWORKING

AWS IOT AND  
AWS GREENGRASS



# 本セッションの紹介範囲





# 本セッションでお伝えする内容

Amazon AI 入門として、  
Amazon AI サービス(Rekognition、Polly、Lex)の

- サービスと機能の概要
- ユースケース
- 利用料金と提供リージョン





# Amazon AI

深層学習に基づくAIサービス

---



## Rekognition

画像分析による  
顔/物体/シーンの認識



## Polly

文章からリアルな  
音声への変換



## Lex

音声や文章を  
使用した会話



# Amazon AI

深層学習に基づくAIサービス

---



## Rekognition

画像分析による  
顔/物体/シーンの認識



## Polly

文章からリアルな  
音声への変換



## Lex

音声や文章を  
使用した会話

# Amazon Rekognition

深層学習を利用した完全マネージド型の画像認識サービス



リアルタイム  
& バッチ  
による画像分析



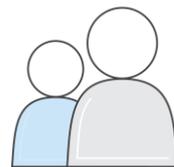
物体および  
シーンの検出



顔の認識



顔の分析



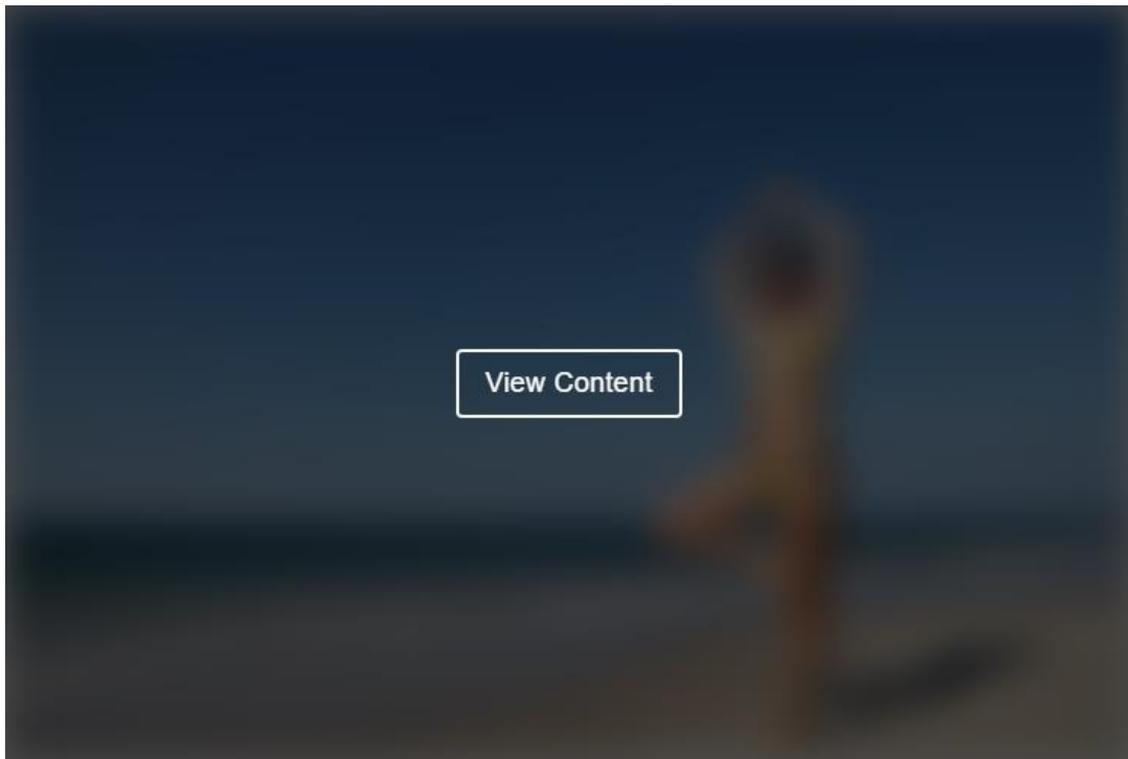
顔の比較

# 物体およびシーンの検出



ラベル	信頼スコア
Bay	99.18%
Beach	99.18%
Coast	99.18%
Outdoors	99.18%
Sea	99.18%
Water	99.18%
Palm_tree	99.21%
Plant	99.21%
Tree	99.21%
Summer	58.3%
Landscape	51.84%
Nature	51.84%
Hotel	51.24%

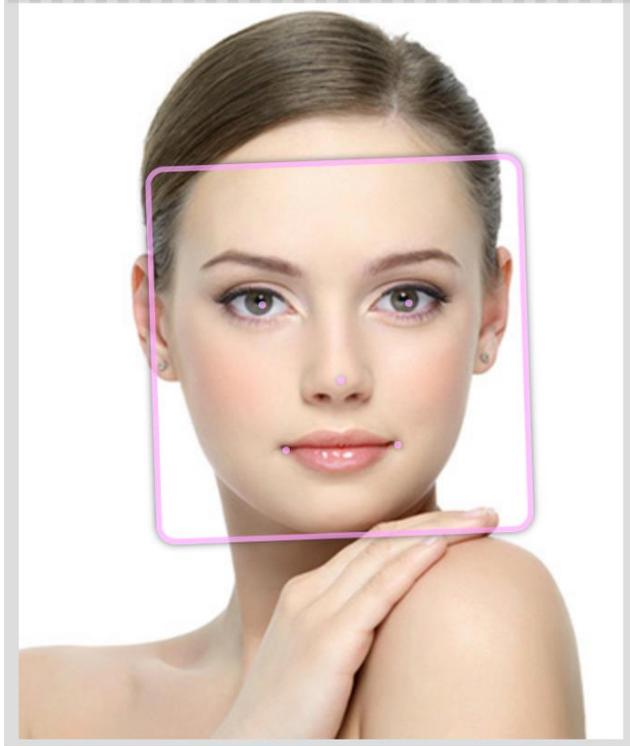
# 画像の節度



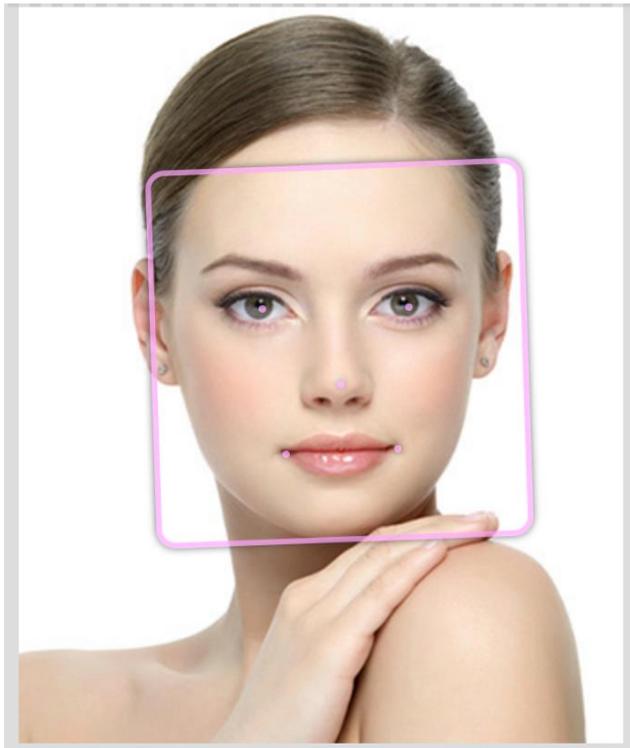
Suggestive	83.5%
Female Swimwear Or Underwear	83.5%

- **明示的なヌード - Explicit Nudity**
  - ヌード
  - 男性のヌード画像
  - 女性のヌード画像
  - 性的な行為
  - 部分的なヌード
- **暗示的 - Suggestive**
  - 女性の水着または下着
  - 男性の水着または下着
  - 露出の多い衣服

# 顔の認識



# 顔の分析



**Gender: Female: 100%**

**Age Range: 26-43 years old**

**Emotion: calm: 73%**

**Eye closed: open (value: 0)**

**Glasses: no glass (value: 0)**

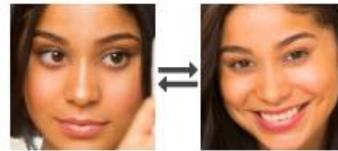
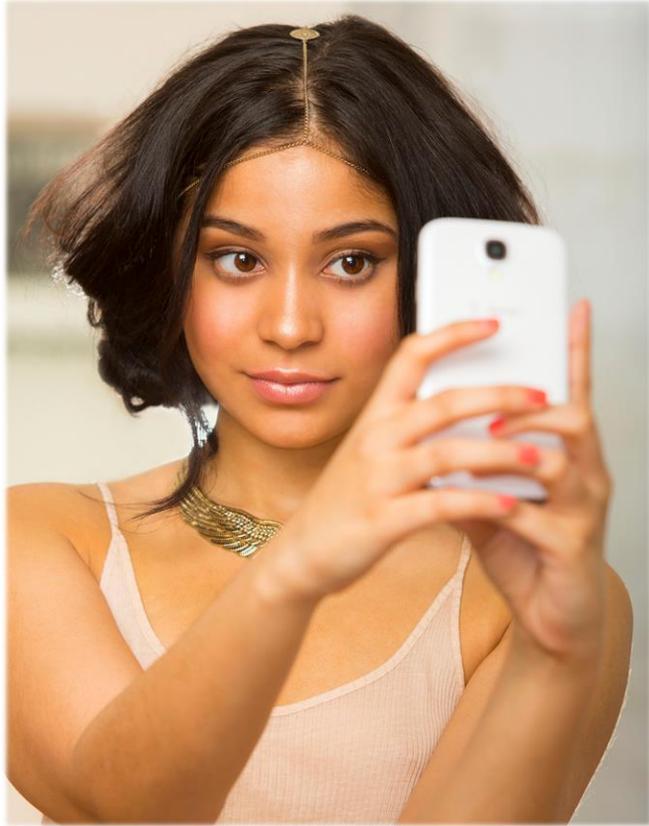
**Sunglasses: false (value: 0)**

**Mouth open: false (value: 0)**

**Mustache: false (value: 0)**

**Beard: no (value: 0)**

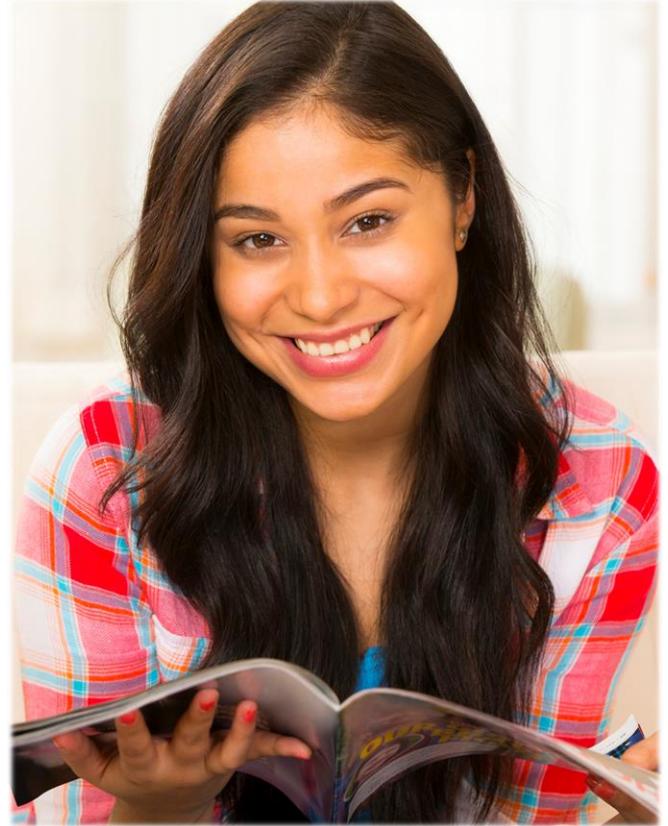
# 顔の比較



Similarity 94%



類似性スコア



# 顔コレクションの作成

顔の特徴をベクトル化した検索可能なインデックス



**IndexFaces**

**SearchFacesByImage**

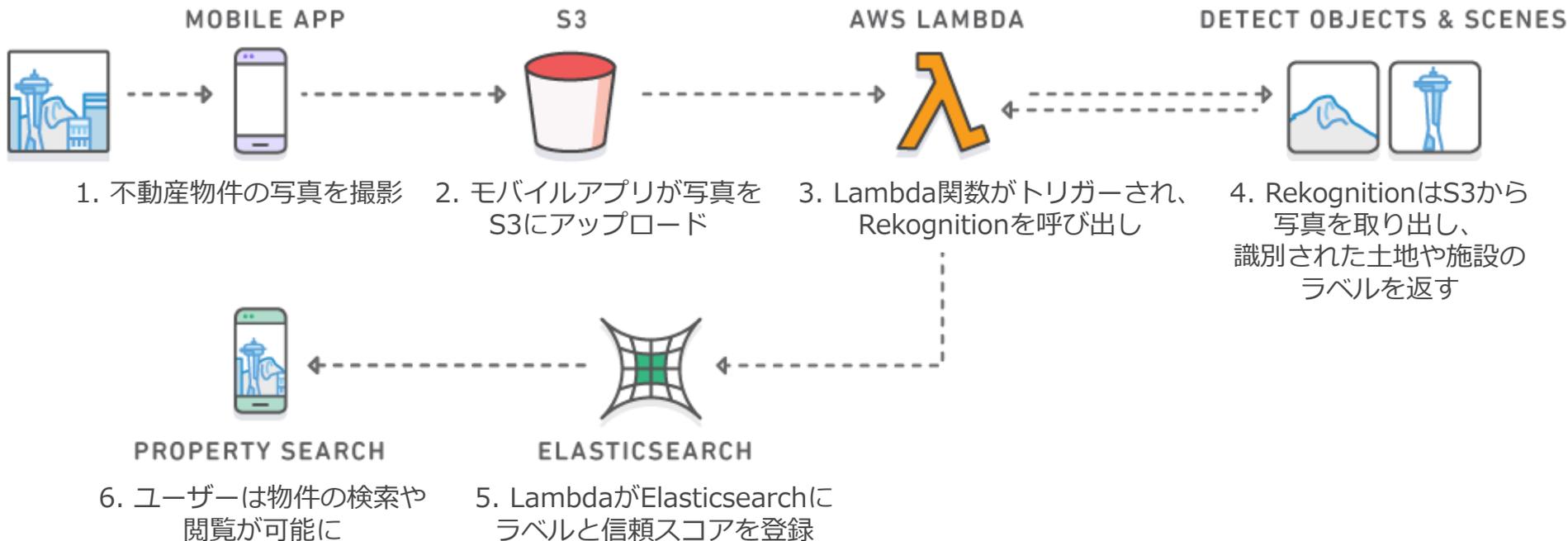
インデックス

コレクション

検索

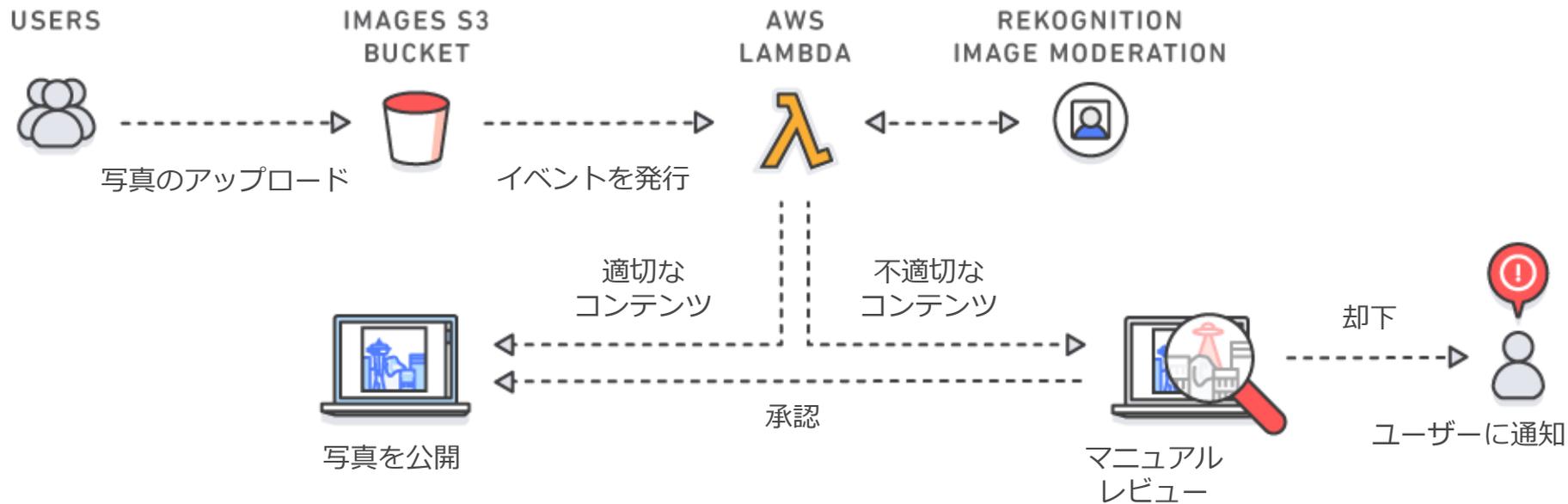
# ユースケース – 検索可能な画像ライブラリ

## 不動産物件の検索



# ユースケース – 画像の節度

ユーザーがアップロードした画像の節度の判定



# ユースケース - 顔ベースのユーザー認証

## 従業員のバッジスキャン

IMAGE CAPTURE



AUTHENTICATED USER

ユーザー認証

APPLICATION

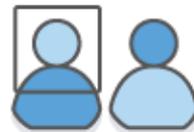


1. 社員証をスキャンする際にアプリケーションがライブ画像を撮影
4. 類似性スコアが80%以上ならOKを返し、そうでない場合はアラートを通知して警備員に検査を促す

ライブ画像 + 社員証画像 S3 URL

信頼スコア(99%)

COMPARE FACES



3. Rekognitionは画像を比較し、類似度を返す

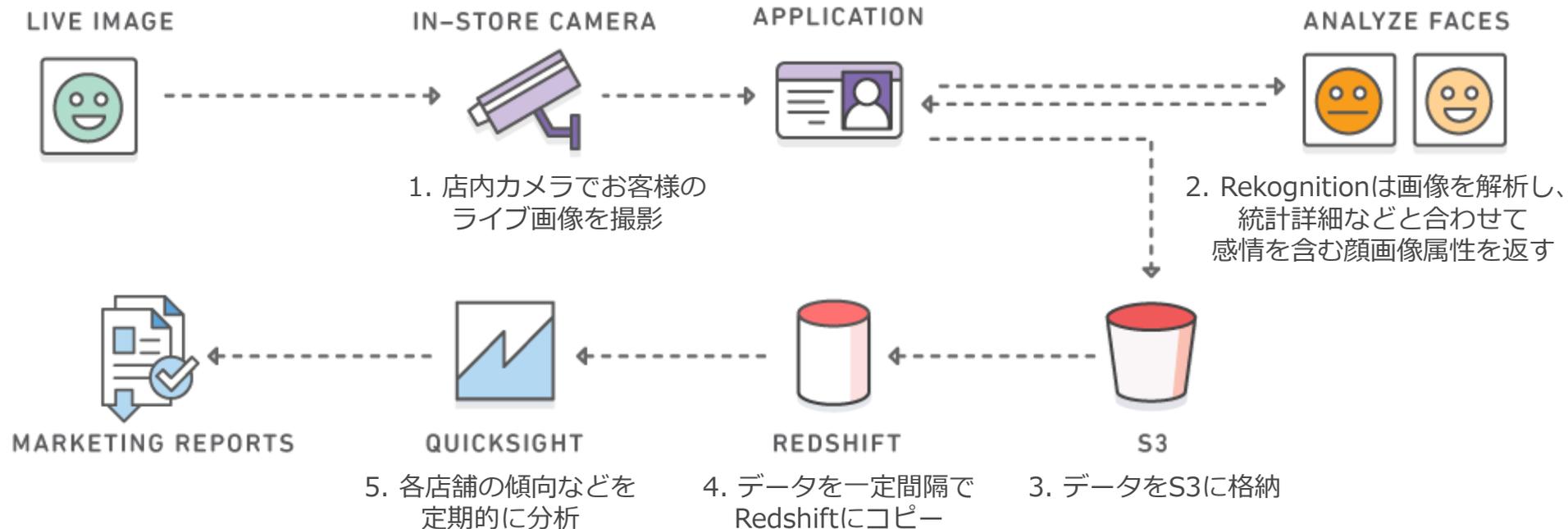


S3

2. S3の社員証画像を取り出し

# ユースケース – 感情分析

## 小売店の感情分析



# ユースケース - 顔認識

## 友人の画像を見つける

REFERENCE IMAGE



SEARCH FACES BY IMAGE



FACE COLLECTION



1. Rekognitionは顔コレクションを検索し、  
類似度の順に一致する可能性のある顔のメタデータを返す

PERSONAL DETAILS TABLE

#0123	#0123
5426	128762
78426	45871
286546	26751
3861	945



END USER



PHOTO APP



S3

4. 類似する写真を閲覧

3. 写真アプリが検索結果を表示

2. 元画像が要求された場合は  
S3から取り出し

# 利用料金と提供リージョン

- Rekognition料金(無料利用枠を超えた場合)

画像分析枠	処理画像1,000枚あたりの料金
1か月あたり画像処理* 100万枚まで	1.00USD
1か月あたり画像処理* 100万枚超1,000万枚まで	0.80USD
1か月あたり画像処理* 1,000万枚超1億枚まで	0.60USD
1か月あたり画像処理* 1億枚超	0.40USD

顔メタデータストレージ	
1か月あたりに保存される顔メタデータ1,000件	0.01USD

\* 各APIで1枚以上の入力画像を受信した場合に、画像処理1枚としてカウント

- 無料利用枠: 最初の1年間は、1か月あたり5,000枚の画像分析を行い、毎月1,000件の顔メタデータを保存可能
- サービス提供リージョン:  
米国東部(バージニア北部)、米国西部(オレゴン)、欧州(アイルランド)

# Amazon Rekognition

深層学習を利用した完全マネージド型の画像認識サービス



高品質



高機能



使いやすく、  
連携も容易



本番利用



低コスト  
(無料利用枠)

<https://aws.amazon.com/rekognition>



# Amazon AI

深層学習に基づくAIサービス

---



## Rekognition

画像分析による  
顔/物体/シーンの認識



## Polly

文章からリアルな  
音声への変換



## Lex

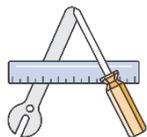
音声や文章を  
使用した会話

# Amazon Polly

深層学習を利用してテキストを生きた話し声に変換するサービス



文章をリアルな  
音声に変換



完全マネージド



24種類の言語



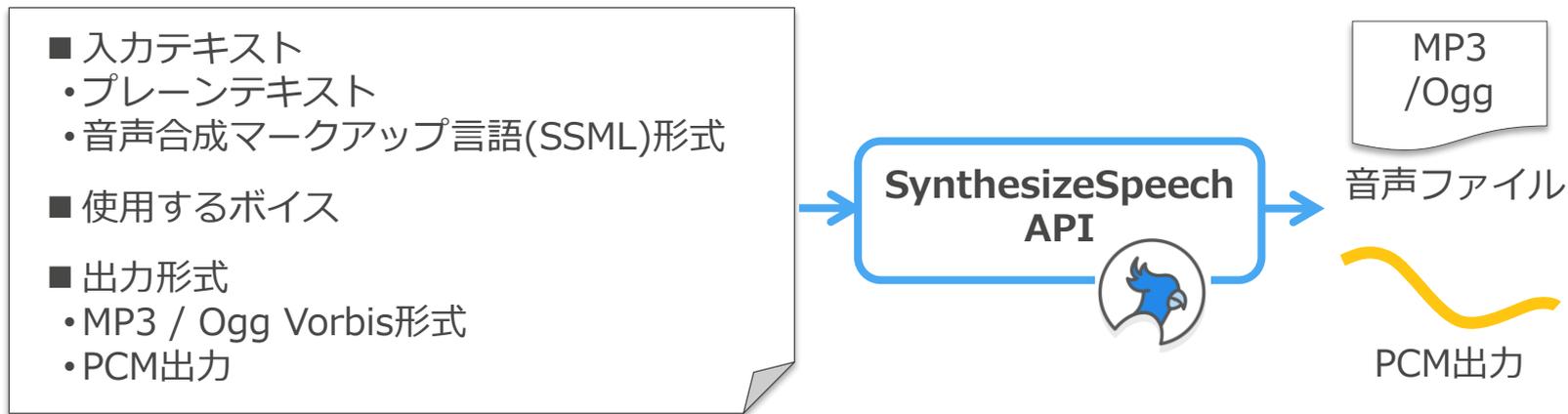
48種類の音声



低レイテンシー、  
リアルタイム

# Text-to-Speech処理

入力されたテキストを音声ストリームに変換



## テキスト読み上げ機能

音声の聴き取り、カスタマイズ、ダウンロード。準備が整ったら統合。

ウィンドウにテキストを入力するか貼り付けて、言語と地域を選択し、音声を選択します。次に、[音声を聴く]を選択し、アプリケーションとサービスに統合します。

プレーンテキスト

SSML



今日は、AWS Summit Tokyo 2017にお越しいただき、ありがとうございます

# 実際に聞いてみましょう



デフォルトのテキストを表示

テキストを消去

▶ 音声を聴く

↓ MP3 のダウンロード

ファイル形式を変更

▶ 発音のカスタマイズ

# 音声品質

- 自然に聞こえる音声
- テキスト変換処理の正確さ
  - 略語、頭字語の展開、日付/時刻の変換、同形異義語の読み分け

Today in Las Vegas, **NV** it's 90°F.



"We **live** for the music", **live** from the Madison Square Garden.



# 発音

- 発音のカスタマイズ
  - 音声合成マークアップ言語(SSML)を使用して、発音、ボリューム、ピッチ、話す速度など、音声のさまざまな面を制御

```
<speak>
```

```
My name is Kawahara. It is spelled
```

```
<prosody rate='x-slow'>
```

```
<say-as interpret-as="characters">Kawahara</say-as>
```

```
</prosody>
```

```
</speak>
```



- エイリアスを定義したレキシコンの適用

*W3C is a Consortium.*

```
<lexeme>
```

```
<grapheme>W3C</grapheme>
```

```
<alias>World Wide Web Consortium</alias>
```

```
</lexeme>
```



# スピーチマーク

開発者が映像体験と会話の同期を可能とするメタデータ

Specify type of file you want to save on your disk

## File format

- MP3
- OGG
- PCM
- Speech Marks

## Sample rate

- 22050Hz
- 16000Hz
- 8000Hz

## Speech Marks Types ?

- Viseme
- Word
- Sentence
- SSML

Cancel

Change

文：1文要素

語句：1単語要素

ビゼーム：音に対応する唇の形

SSML：<mark>タグ

170419063128084.marks - Notepad

```
{ "time": 0, "type": "sentence", "start": 7, "end": 10, "value": "Hi!" }
{ "time": 6, "type": "word", "start": 7, "end": 9, "value": "Hi" }
{ "time": 421, "type": "sentence", "start": 11, "end": 27, "value": "My name is Tara." }
{ "time": 808, "type": "word", "start": 11, "end": 13, "value": "My" }
{ "time": 1025, "type": "word", "start": 14, "end": 18, "value": "name" }
{ "time": 1476, "type": "word", "start": 19, "end": 21, "value": "is" }
{ "time": 1671, "type": "word", "start": 22, "end": 26, "value": "Tara" }
{ "time": 2112, "type": "sentence", "start": 28, "end": 76, "value": "I am excited to talk about Polly's new features." }
{ "time": 2498, "type": "word", "start": 28, "end": 29, "value": "I" }
{ "time": 2696, "type": "word", "start": 30, "end": 32, "value": "am" }
{ "time": 2844, "type": "word", "start": 33, "end": 40, "value": "excited" }
{ "time": 3364, "type": "word", "start": 41, "end": 43, "value": "to" }
{ "time": 3503, "type": "word", "start": 44, "end": 48, "value": "talk" }
```

# 24種類の言語で48種類の音声

ヨーロッパ/中東/アフリカ地域:

言語	女性	男性
デンマーク語	Naja	Mads
オランダ語	Lotte	Ruben
フランス語	Céline	Mathieu
ドイツ語	Marlene	Hans
	Vicki	
アイスランド語	Dóra	Karl
イタリア語	Carla	Giorgio
ノルウェー語	Liv	
ポーランド語	Ewa	Jacek
	Maja	Jan
ポルトガル語(イベリア)	Inês	Cristiano
ルーマニア語	Carmen	
ロシア語	Tatyana	Maxim
スペイン語(カスティリヤ)	Conchita	Enrique
スウェーデン語	Astrid	
トルコ語	Filiz	
英語(英国)	Amy	Brian
	Emma	
英語(ウェールズ)		Geraint
ウェールズ語	Gwyneth	

北アメリカ/南アメリカ地域:

言語	女性	男性
フランス語(カナダ)	Chantal	
ポルトガル語(ブラジル)	Vitória	Ricardo
英語(米国)	Joanna	Joey
	Salli	Justin
	Kendra	
	Kimberly	
	Ivy	
スペイン語(米国)	Penélope	Miguel

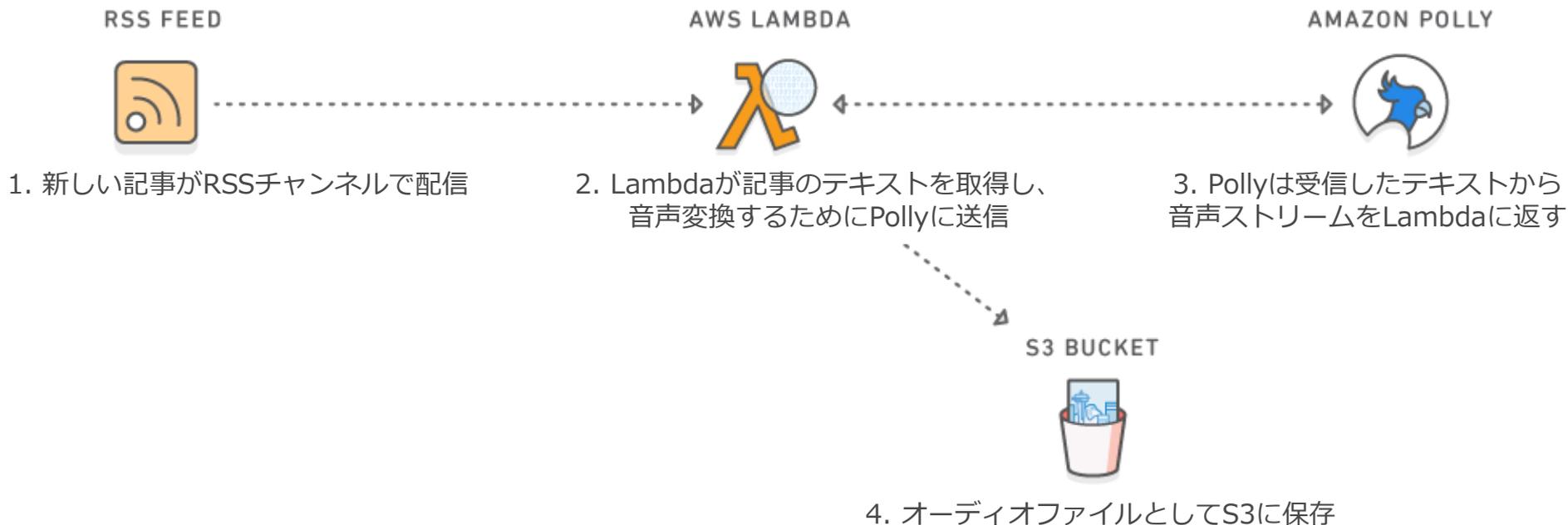
アジア太平洋地域:

言語	女性	男性
英語(オーストラリア)	Naja	Mads
英語(インド)	Raveena	
日本語	Mizuki	



# ユースケース - コンテンツの作成

記事を音声に変換してMP3でダウンロードする



# ユースケース – 教育/Eラーニング

## 音声の再生および発音されるテキストのハイライト

### 1. テキストコンテンツの準備



6. アプリケーションが  
テキストハイライトに合わせて  
音声で読み上げ

### 2. アプリケーションが音声をリクエスト



### 3. Pollyは音声ストリームを返す

### 4. アプリケーションがスピーチマークをリクエスト



### 5. PollyはスピーチマークのJSONストリームを返す



Amazon Polly

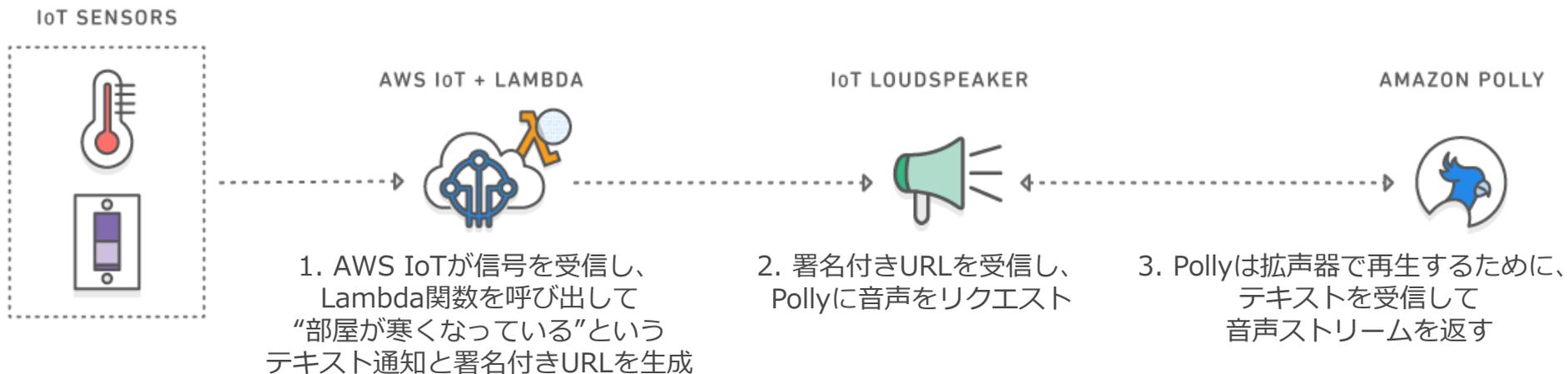
# ユースケース – カスタマーコンタクトセンター

コンタクトセンターは音声による顧客への回答にPollyを使用



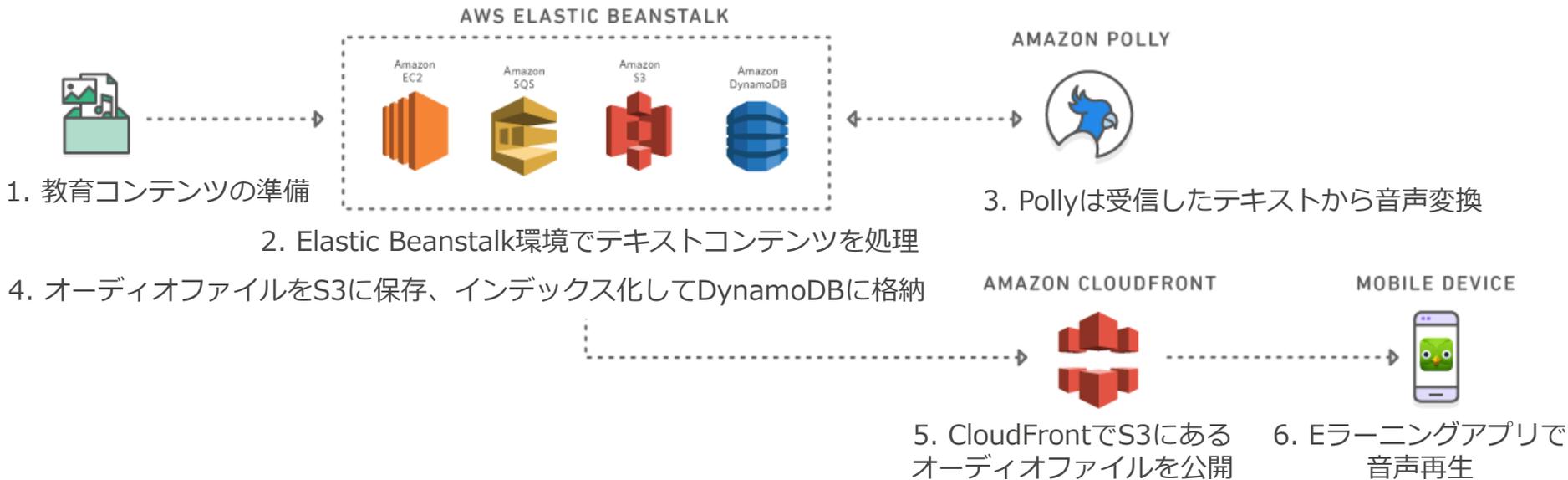
# ユースケース – IoT

## 口頭による温度通知



# ユースケース - 言語学習

新しい言語の話し方を教えるアプリケーションの利便性向上にPollyを使用



# 利用料金と提供リージョン

- Polly料金(無料利用枠を超えた場合):  
音声リクエスト100万文字あたり4.00USDの従量課金制、  
スピーチマークリクエスト100万文字あたり4.00USDの従量課金制
- 無料利用枠:  
最初の1年間は、1か月あたり500万文字まで、  
音声またはスピーチマークリクエストを利用可能
- サービス提供リージョン:  
米国東部(バージニア北部 / オハイオ)、米国西部(オレゴン)、  
欧州(アイルランド)

# Amazon Polly

深層学習を利用してテキストを生きた話し声に変換するサービス



高品質



高機能



使いやすく、  
連携も容易



本番利用



低コスト  
(無料利用枠)

<https://aws.amazon.com/polly>



# Amazon AI

深層学習に基づくAIサービス

---



## Rekognition

画像分析による  
顔/物体/シーンの認識



## Polly

文章からリアルな  
音声への変換



## Lex

音声や文章を  
使用した会話

# Amazon Lex

音声やテキストを使用した会話型インターフェイスを構築するサービス



音声やテキストの  
“チャットボット”



Alexaの  
深層学習技術



モバイルデバイス  
やウェブアプリ  
から音声会話



Slackや  
FBメッセージャーで  
テキスト会話

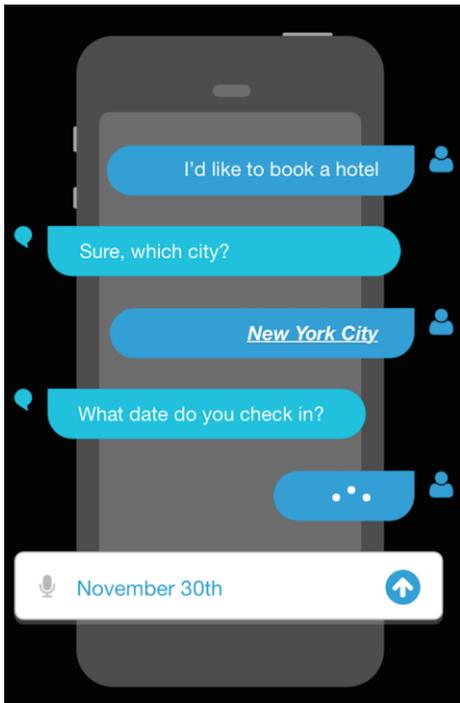


エンタープライズ  
コネクター

Salesforce  
Microsoft Dynamics  
Marketo  
Zendesk  
Quickbooks  
HubSpot

# ボット構造

## BookHotel



## Intents

ユーザーが達成したいゴール

## Utterances

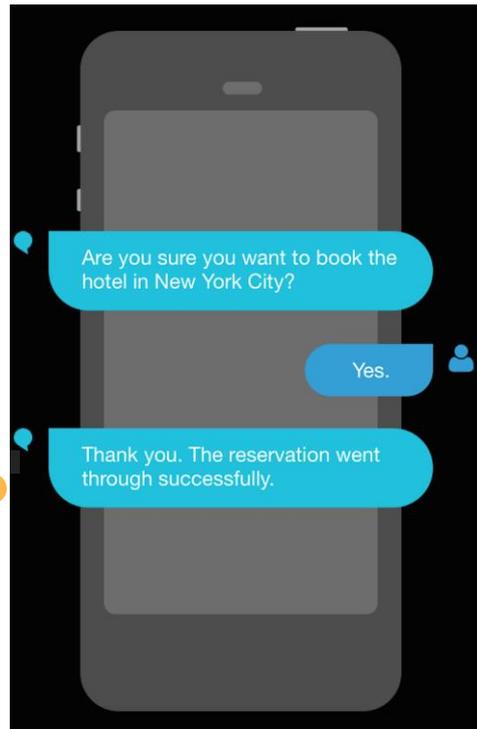
インテントを呼び出すために  
発声あるいは入力されるフレーズ

## Slots

インテントを実現するために必要な  
ユーザーからの入力データ

## Fulfillment

インテントを実行するための  
ビジネスロジック



# “Book a Hotel” 会話の流れ



# 効率的で直観的な開発ツール

< BookTrip Latest ▾

Build

Publish



Editor

Settings

Channels

Monitoring

Intents



BookCar

BookHotel

Slot types



RoomTypeValues

Error Handling

BookHotel 1 ▾

▼ Sample utterances ⓘ

Book a hotel

I want a make hotel reservations

Book a {Nights} night stay in {Location}

▶ Lambda initialization and validation ⓘ

▼ Slots ⓘ

1.



Location

AMAZON.US...

Built-in ▾

2.



CheckInDate

AMAZON.DATE ▾

Built-in ▾

3.



Nights

AMAZON.NU... ▾

Built-in ▾

4.



RoomType

RoomTypeVa... ▾

1 ▾

Test Bot



Build: Latest | Status: READY

book a hotel

What city will you be staying in?

las vegas

What day do you want to check in?

2017/6/2

How many nights will you be staying?

Clear

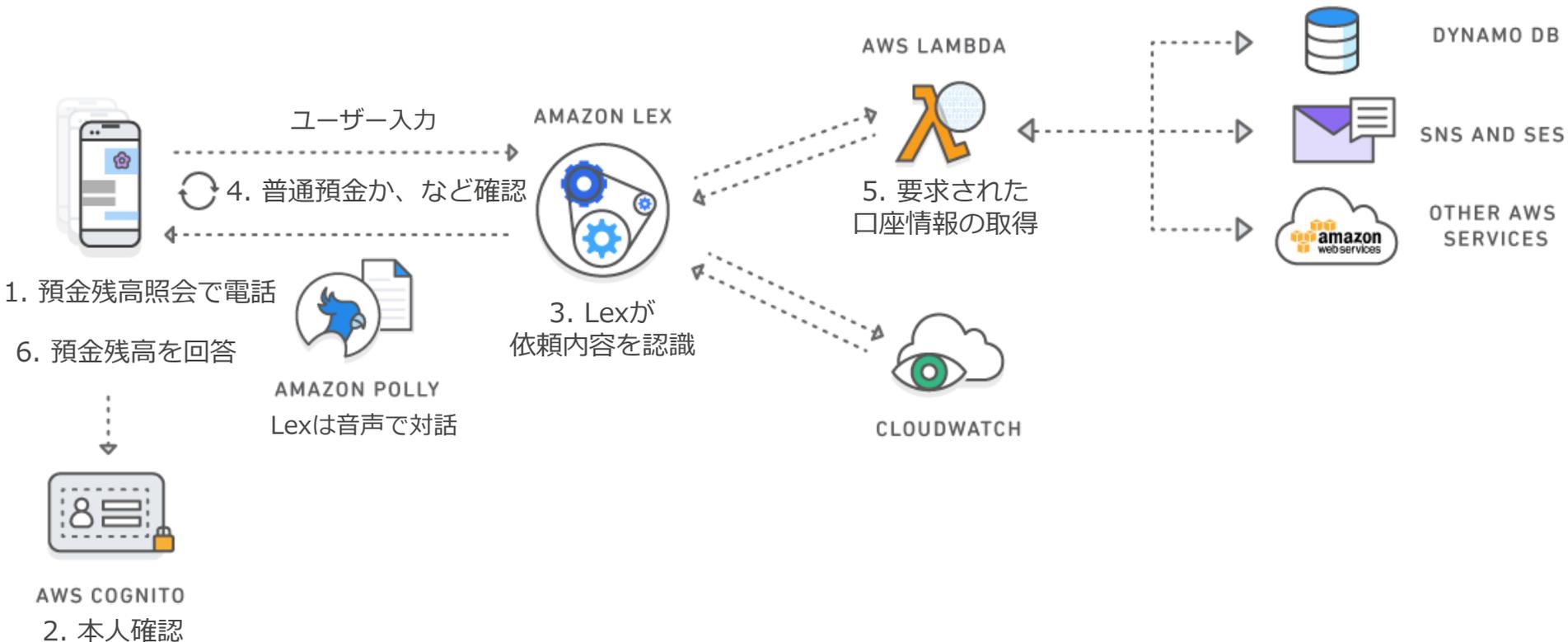


|Type to your bot...



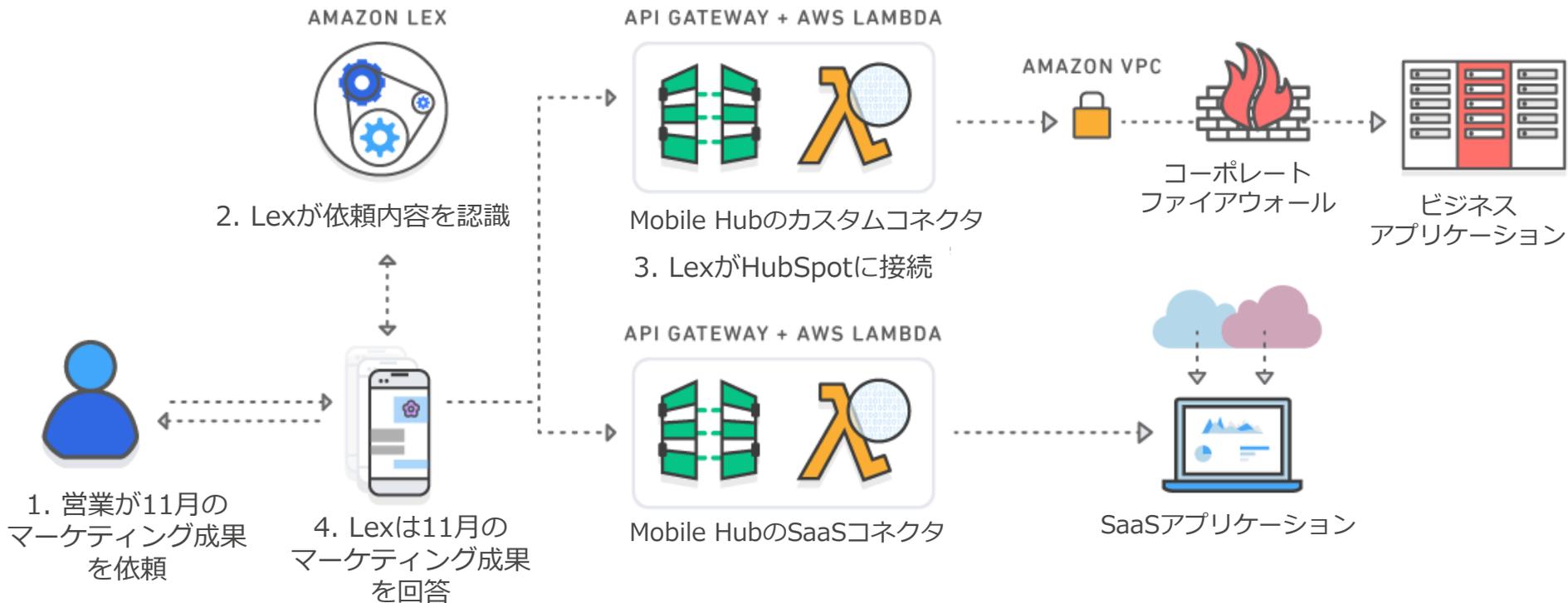
# ユースケース – アプリケーションボット

## Amazon Lexチャットボットから銀行情報を受け取る



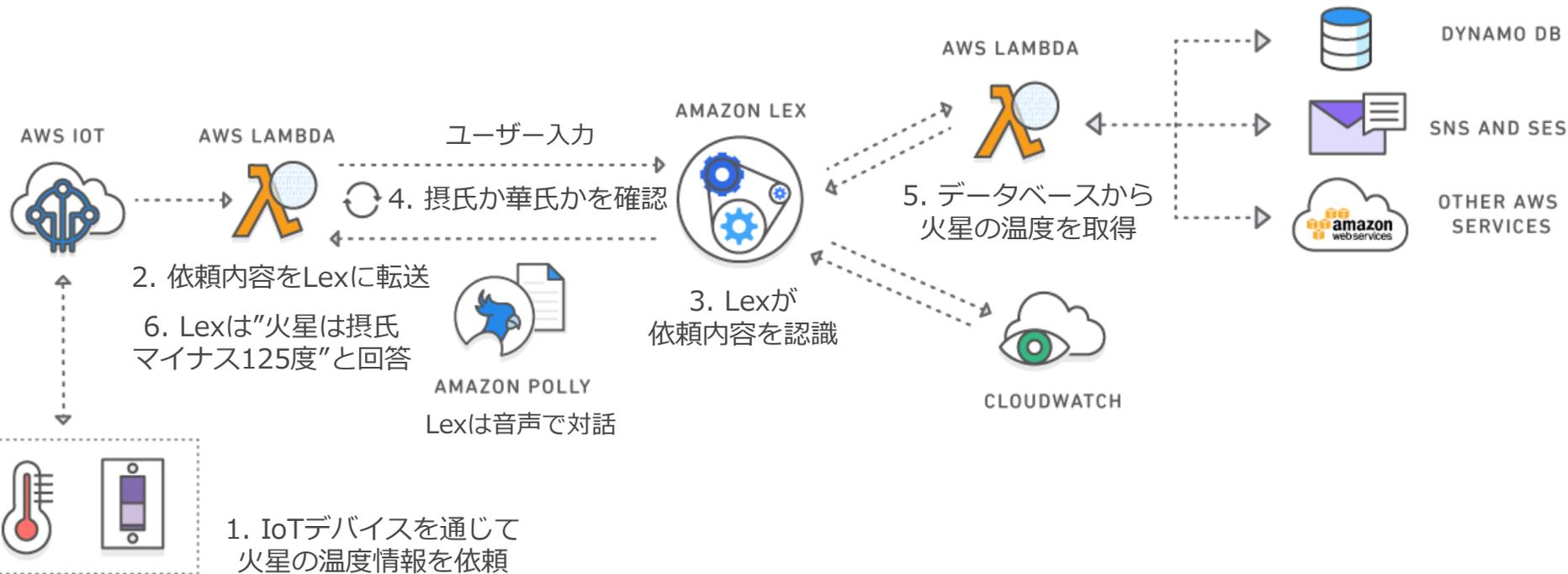
# ユースケース – エンタープライズ生産性ボット

マーケティングデータを受け取るために  
エンタープライズアプリケーションに接続するAmazon Lexボットを構築する



# ユースケース – IoT

教育と探査のための乗り物としてAmazon Lexボットを使用する



# 利用料金と提供リージョン

- Lex料金(無料利用枠を超えた場合):  
音声リクエスト1件あたり0.004USDの従量課金制、  
テキストリクエスト1件あたり0.00075USDの従量課金制
- 無料利用枠:  
最初の1年間は、1か月あたり最大で10,000回のテキストリクエストと  
5,000回の音声リクエストを処理可能
- サービス提供リージョン:  
米国東部(バージニア北部)

# Amazon Lex

音声やテキストを使用した会話型インターフェイスを構築するサービス



高品質



高機能



使いやすく、  
連携も容易



本番利用



低コスト  
(無料利用枠)

<https://aws.amazon.com/lex>



# Amazon AI

深層学習に基づくAIサービス

---



## Rekognition

画像分析による  
顔/物体/シーンの認識



## Polly

文章からリアルな  
音声への変換



## Lex

音声や文章を  
使用した会話

# さあ、Amazon AIを始めましょう

## Amazon AI

強力な人工知能をすべての開発者に

まずは無料で始める



<https://aws.amazon.com/amazon-ai>

AWS

S U M M I T

